基于深度学习的图像风格迁移研究综述 *

陈淑環，韦玉科，徐　乐，董晓华，温坤哲

(广东工业大学 计算机学院，广州 510006)

摘　要：图像风格迁移是一种用不同风格渲染图像语义内容的图像处理方法。随着深度学习的兴起，图像风格迁移获得了进一步的发展，并取得了一系列突破性的研究成果。其出色的风格迁移能力引起了学术界和工业界的广泛关注，具有重要的研究价值。为推进基于深度学习的图像风格迁移的技术研究，对目前的主要方法和代表性工作进行了归纳与探讨。首先回顾了非参数的图像风格迁移，详细介绍了目前主要的基于深度学习的图像风格迁移的基本原理和方法，分析了图像风格迁移在相关领域中的应用前景，最后总结了基于深度学习的图像风格迁移目前存在的问题与未来的研究方向。

关键词：图像风格迁移；深度学习；迁移学习；纹理合成

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2018.05.0270

Survey of image style transfer based on deep learning

Chen Shuhuan, Wei Yuke, Xu Le, Dong Xiaohua, Wen Kunzhe

(*School of Computer Science Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: Image style transfer is an image processing method that renders the semantic content of an image in different styles. With the rise of deep learning, image style transfer has gained further development and has achieved a series of breakthrough research results. Its outstanding style transfer ability has attracted wide attention from academic and industrial circles and has important research value. In order to promote the technology research of image style transfer based on deep learning, this paper summarizes and discusses the current major methods and representative work. Firstly, this paper reviews the non-parametric image style transfer, and introduces the basic principles and methods of image style transfer based on deep learning, and analyzes the application prospect of image style transfer technology in related fields. At last, this paper summarizes the existing problems and future research directions of image style transfer based on deep learning.

Key words: image style transfer; deep learning; transfer learning; texture synthesis

0　引言

传统非参数的图像风格迁移方法主要基于物理模型的绘制和纹理的合成。Efros 等人[错误!未找到引用源。]提出了一种简单的纹理算法，通过对样本纹理进行拼接和重组以合成新的纹理。Hertzmann 等人[错误!未找到引用源。]提出了一种基于类推思想的方法，通过图像特征映射关系合成具有新纹理的图像。张海嵩等人[错误!未找到引用源。]运用多层纹理阵列、国画光照模型、提取轮廓线等模块，实时绘制 3D 中国画效果的山峦场景。钱小燕等人[错误!未找到引用源。]提出了一种邻域一致性度量方法，通过把统计特性引入相似性度量中，以提高图像匹配点搜索的效率。虽然这些方法已经获得了可观的效果，但是非参数的图像风格迁移方法只能提取图像的底层特征，而非高层抽象特征，在处理颜色和纹理较复杂的图像时，最终的图像合成效果较为粗糙，难以符合实际需求。

随着深度学习[错误!未找到引用源。错误!未找到引用源。错误!未找到引用源。]的兴起，Gatys 等人[错误!未找到引用源。]开创性地提出了一种基于卷积神经网络的图像风格迁移，发现了可以使用卷积神经网络来将图像的内容抽象特征表示和风格抽象特征表示进行分离，并通过独立处理这些高层抽象特征表示来有效地实现图像风格迁移，获得了非常可观的艺术效果，如图 1 所示。该算法的核心思想是使用预训练 VGG 模型[错误!未找到引用源。]分别提取内容图像和风格图像的高层抽象特征表示，然后从随机噪声图像开始，通过迭代优化方式生成具有原内容和新风格的合成图像。

Gatys 等人[错误!未找到引用源。]的工作引起了学术界和工业界的广泛关注。在学术界，大量的后续研究被提出，主要包括基于图像迭代和基于模型迭代两个方面。其中，根据图像风格获取方式

收稿日期：2018-05-02；**修回日期：**2018-06-19　**基金项目：**广东省省级科技计划资助项目（2014A040402007）

作者简介：陈淑環（1994-），女，江西上饶人，硕士，主要研究方向为计算机视觉、深度学习（cxshuan@126.com）；韦玉科（1965-），女，副教授，博士，主要研究方向为人工智能和计算机测控技术；徐乐（1994-），女，江西赣州人，硕士，主要研究方向为图像处理；董晓华（1989-），男，广东韶关人，硕士，主要研究方向为图像处理；温坤哲（1993-），男，广东汕尾人，硕士，主要研究方向为深度学习．

的不同，基于图像迭代的方法可以归纳为基于最大均值差异（MMD, Maximum Mean Discrepancy）[8][14]、基于马尔可夫随机场（MRF, Markov Random Field）[15]、以及基于深度图像类比（DIA, Deep Image Analogy）[16]。而根据模型迭代方式的不同，基于模型迭代的主要方法可以归纳为基于生成模型[17][18][19]和基于图像重构解码器[20][21]。这些方法成功地应用于工业应用软件中，如 Prisma, Ostagram，Deep Forger 等热门应用软件。

内容图 风格图 效果图

图 1　Gatys 等人的图像风格迁移效果

本文系统综述了目前主要的基于深度学习的图像风格迁移方法，分析了其最新的研究成果和应用前景，以及对目前尚待解决的问题展开了深入探讨，并提出了一些具有实际参考价值的建议，为进一步的深入研究奠定了一定的基础，最后总结了未来所面临的挑战和发展趋势。

1　基于深度学习的图像风格迁移方法

本节描述了目前主要的基于深度学习的图像风格迁移方法，包括基于图像迭代和基于模型迭代两类，如表 1 所示。第一类是直接在白噪声图像上进行优化迭代来进行风格迁移，其优化目标是白噪声图像，而第二类是迭代地优化神经网络模型，以网络前馈的方式实现快速风格迁移，其优化目标是神经网络模型。以下将对这两类方法展开详细的探讨。

1.1 基于图像迭代的图像风格迁移方法

基于图像迭代的目标是使得白噪声图像同时匹配内容图像的内容特征表示和风格图像的风格特征表示，最终获得风格化的合成图像。以下将对基于最大均值差异、基于马尔可夫随机场和基于深度图像类比这三类具有代表性的方法展开详细的探讨。

1.1.1 基于最大均值差异

Gatys 等人[错误!未找到引用源。]最早发现通过重建 VGG 网络中间层的抽象特征表示，能够从任意图像中提取抽象的内容表示，而通过构造 Gram 矩阵[错误!未找到引用源。]可以提取出任意图像的风格特征表示。Li 等人[错误!未找到引用源。]从理论上证明了 Gram 矩阵的匹配方式等价于最小化特定的最大均值差异。因此，本文将基于 Gram 矩阵的风格迁移方法归类为基于最大均值差异的方法。具体而言，,给定的白噪声图像 x，内容图像 x_c 和风格图像 x_s，Gatys 等人[错误!未找到引用源。]的方法的总损失函数表示如下：

$$L_t(x,x_c,x_s)=\alpha L_c(x,x_c)+\beta L_s(x,x_s) \tag{1}$$

其中: α 表示图像内容损失函数 $L_c(x,x_c)$ 的权重系数，β 表示图像风格损失函数 $L_s(x,x_s)$ 的权重系数。而图像内容损失函数 L_c 表示如下：

$$L_c(x_c,x)=\frac{1}{2}\sum_{i=1}^{N_l}\sum_{j=1}^{M_l}\left(F_{ij}^l-P_{ij}^l\right)^2 \tag{2}$$

其中: F^l 表示白噪声图像在 VGG 模型中第 l 层的内容特征表示，F_{ij}^l 表示第 l 层中第 i 个过滤器上位置 j 处的激活值，P^l 表示内容图像在 VGG 中第 l 层的内容特征表示。而风格总损失函数 L_s 表示如下：

$$L_s\left(x,x_s\right)=\sum_{l=0}^{L}\omega_l E_l \tag{3}$$

其中: L 表示在 VGG 网络中用来提取图像风格特征表示的卷积层的总层数，ω_l 表示对应卷积层 l 的图像风格损失函数的权重因子，E_l 表示在 l 层中的风格损失函数：

$$E_l=\frac{1}{4N_l^2M_l^2}\sum_{i=1}^{N_l}\sum_{j=1}^{M_l}\left(G_{ij}^l-A_{ij}^l\right)^2 \tag{4}$$

其中: G_{ij}^l 和 A_{ij}^l 分别表示内容图像和风格图像的 Gram 矩阵，N_l 表示 l 层中滤波器的数量，M_l 表示 l 层中的特征图的大小。Gatys 等人使用 Gram 矩阵来定义风格损失函数，Gram 矩阵 G_{ij}^l 表示如下：

$$G_{ij}^{l}=\sum_{k}F_{ik}^{l}F_{jk}^{l} \tag{5}$$

其中：G_{ij}^{l} 为在层 l 中第 i 个和第 j 个矢量化特征映射之间的内积，F_{ik}^{l} 表示图像在第 l 层中第 i 个过滤器上位置 k 处的激活值。

表 1　图像风格迁移方法归纳表

类别	基本方法	代表性工作	优点	缺点
基于图像迭代	最大均值差异	Gatys 等人**错误!未找到引用源。**。	合成图像的质量高、可控性好，易于调参，无需训练数据。	计算时间较长，对预训练模型的依赖性大。
	马尔可夫随机场	Li 等人**错误!未找到引用源。**。		
	深度图像类比	Liao 等人**错误!未找到引用源。**。		
基于模型迭代	生成模型	Johnson 等人**错误!未找到引用源。**，Zhu 等人**错误!未找到引用源。**。	计算速度快，可用于视频快速风格化，目前工业应用软件的主流技术。	图像生成质量有待进一步提高，需要大量的训练数据。
	图像重构解码器	Li 等人**错误!未找到引用源。**。		

随后，Nikulin 等人[错误!未找到引用源。]深入探索了 Gatys 等人[错误!未找到引用源。]方法的原理，讨论了不同的超参数和程式化属性对图像风格迁移效果的影响。Novak 等人[错误!未找到引用源。]提出了改进 Gatys 等人[错误!未找到引用源。]方法的一些途径。为了更好地解释和完善 Gram 矩阵，Li 等人[错误!未找到引用源。]在 Gatys 等人[错误!未找到引用源。]工作的基础上对 Gram 矩阵展开了深入的探讨，提出了使用不同的核函数来改进风格损失函数，如线性核函数和高斯核函数等。另外，Gatys 等人[错误!未找到引用源。]方法在迭代优化期间不稳定，容易影响纹理合成，为了改善这个问题，Risser 等人[错误!未找到引用源。]引入了直方图损失函数，解决了因迭代优化过程不稳定而产生图像纹理错乱的问题。Yin [错误!未找到引用源。]提出了基于内容感知的方法，该方法能有效地控制图像内容和图像纹理的合成，进一步提高合成图像的分辨率。

1.1.2 基于马尔可夫随机场

马尔可夫随机场是非参数图像合成的经典框架[错误!未找到引用源。]，它描述了具有同类特征信息的集合。Li 等人[错误!未找到引用源。]最早提出了马尔可夫随机场与深度卷积神经网络结合的方法，将图像特征映射分割成许多区域块并进行匹配，以提高合成图像在视觉上的合理性。具体而言，给定内容图像 x_c 和风格图像 x_s，设合成目标图像为 x，问题求解可以表示如下:

$$\begin{aligned}x^{*}=&\arg\min_{x}E_{s}(\varphi(x),\varphi(x_{s}))\\&+\alpha_{1}E_{c}(\varphi(x),\varphi(x_{c}))\\&+\alpha_{2}\gamma(x)\end{aligned} \tag{6}$$

其中：E_s 表示风格损失函数，E_c 表示内容损失函数，$\gamma(x)$ 表示用来平滑合成图像的正则化选项，$\varphi(x)$ 表示神经网络模型层次中的特征映射集合，α_1 表示内容图像损失函数的权重系数，α_2 表示正则化选项的权重系数。图像风格损失函数 E_s 的表达式如下：

$$E_{s}(\varphi(x),\varphi(x_{s}))=\sum_{i=1}^{m}\left\|\psi_{i}(\varphi(x))-\psi_{NN(i)}(\varphi(x_{s}))\right\|^{2} \tag{7}$$

其中：m 是 $\psi(\varphi(x))$ 的基数，即区域块的数量，$\psi_i(\varphi(x))$ 表示 $\varphi(x)$ 中的一个区域块，而对于每一个区域块 $\psi_i(\varphi(x))$ 是通过归一化互相关的方法查找其最佳匹配块 $\psi_{NN(i)}(\varphi(x_s))$，表达式如下：

$$NN(i):=\arg\min_{j=1,\ldots,m_{s}}\frac{\psi_{i}(\varphi(x))\cdot\psi_{j}(\varphi(x_{s}))}{|\psi_{i}(\varphi(x))|\cdot|\psi_{j}(\varphi(x_{s}))|} \tag{8}$$

图像内容损失函数 E_c 的表达式如下：

$$E_{c}(\varphi(x),\varphi(x_{c}))=\|\varphi(x)-\varphi(x_{c})\|^{2} \tag{9}$$

正则化选项 $\gamma(x)$ 的表达式如下：

$$\gamma(x)=\sum_{i,j}\left(\left(x_{i,j+1}-x_{i,j}\right)^{2}+\left(x_{i+1,j}-x_{i,j}\right)^{2}\right) \tag{10}$$

随后，Champandard 等人[错误!未找到引用源。]在 Li 等人工作的基础上加入了手工添加的图像语义映射，增强了对合成结果的控制，使得合成结果的结构更为合理，并且显著提高了合成图像的质量。

1.1.3 基于深度图像类比

图像类比的概念最初由 Hertzmann 等人[错误!未找到引用源。]提出，主要处理的问题是深入挖掘图像之间的映射关系。为了更好地在两个输入图像之间找到语义上有意义的密集对应关系，Liao 等人[错误!未找到引用源。]将图像类比的概念与深度学习进行结合，提出了一种通过区域块匹配迭代优化的深度图像类比方法。这种方法可以将图像类比的概念应用到深层网络特征空间中，寻找语义上有意义的稠密对应关系，从而提高图像风格迁移的有效性。

深度图像类比的映射关系可以表示为 A:A*::B:B*，其中，A*和 B 为未知变量。在这个映射关系中具有两个约束条件：A 和 A*或 B 和 B*具有相似的图像内容特征；A 和 B 或 A*和 B*

具有相似的图像风格特征。因此，图像映射关系可表示为 $A(p)=B(\varphi_{a\to b}(p))$ 和 $A^*(p)=B^*(\varphi_{a\to b}(p))$。而 $\varphi_{b\to a}$ 表示逆映射，则有 $\varphi_{b\to a}(\varphi_{a\to b}(p))=p$ 和 $\varphi_{a\to b}(\varphi_{b\to a}(p))=p$ 。

深度图像类比方法首先使用预训练 VGG 模型分别计算出已知图像 A 和 B*的各层抽象特征表示 F_A^L 和 $F_{B^*}^L$，其中，L 表示使用预训练 VGG 模型中的层数。在第 L 层中可以通过最近邻域搜索（nearest neighbor field search, NNFS）求得深度图像类比的映射关系。虽然 A*和 B 为未知变量，但是根据类比映射关系的第一个约束，可以认为 A 和 A*或 B 和 B*在预训练 VGG 模型中具有相似的高层抽象特征表示，因此可以有 $F_{A^*}^5=F_A^5$ 和 $F_{B^*}^5=F_B^5$。那么，$\varphi_{a\to b}^L(p)$ 映射关系函数可以表示如下：

$$\begin{aligned}\varphi_{a\to b}^L(p)=\arg\min_q &\sum_{x\in N(p),y\in N(q)}\left\|F_A^L(x)-F_B^L(y)\right\|^2\\ &+\sum_{x\in N(p),y\in N(q)}\left\|F_{A^*}^L(x)-F_{B^*}^L(y)\right\|^2\end{aligned}\tag{11}$$

其中：点 $N(p)$ 表示点 p 附近的区域块，同理可得 $\varphi_{b\to a}^L(p)$ 映射关系函数。根据以上深度图像类比的映射关系，在预训练 VGG 模型的抽象特征空间中，通过卷积映射函数 $CNN_{L-1}^L(\cdot)$ 和逆映射特征 $R_{B^*}^{L-1}$ 的反卷积操作，由高层往低层迭代优化，最终可求得具有 A 内容和 B*风格的合成图像 A*，以及具有 B*内容和 A*风格的合成图像 B。具体算法过程可以用伪代码描述如下所示。

算法：深度图像类比算法

输入：两张 RGB 颜色空间图像 A 和 B^*。

输出：两个像素空间映射关系函数 $\varphi_{a\to b}$ 和 $\varphi_{b\to a}$；以及两张 RGB 颜色空间图像 A^* 和 B。

运算过程：

$\{F_A^L\}_{L=1}^5$，$\{F_{B^*}^L\}_{L=1}^5$ ←输入 A 和 B^* 到预训练 VGG 模型中提取抽象特征。

$F_{A^*}^5=F_A^5$，$F_B^5=F_{B^*}^5$，随机初始化映射关系函数 $\varphi_{a\to b}$ 和 $\varphi_{b\to a}$。

for L=5 to 1 do:

  最近邻域搜索：

    $\varphi_{a\to b}^L$ ← map F_A^L to F_B^L，$F_{A^*}^L$ to $F_{B^*}^L$，

    $\varphi_{b\to a}^L$ ← map F_B^L to F_A^L，$F_{B^*}^L$ to $F_{A^*}^L$.

  If L>1 then:

    图像重构：

      使用函数 $\varphi_{a\to b}^L$ 将 $F_{B^*}^L$ 逆转到 $F_{B^*}^L(\varphi_{a\to b}^L)$。

    使用 $F_{B^*}^L(\varphi_{a\to b}^L)$ 和 $CNN_{L-1}^L(\cdot)$ 对 $R_{B^*}^{L-1}$ 进行反卷积。

    $F_{A^*}^{L-1}$ ←加权映射 F_A^{L-1} 和 $R_{B^*}^{L-1}$。

    使用函数 $\varphi_{b\to a}^L$ 将 F_A^L 逆转到 $F_A^L(\varphi_{b\to a}^L)$。

    使用 $F_A^L(\varphi_{b\to a}^L)$ 和 $CNN_{L-1}^L(\cdot)$ 对 R_A^{L-1} 进行反卷积。

    F_B^{L-1} ←加权映射 $F_{B^*}^{L-1}$ 和 R_A^{L-1}。

    最近邻搜索上采样：

    上采样 $\varphi_{a\to b}^L$ 到 $\varphi_{a\to b}^{L-1}$。

    上采样 $\varphi_{b\to a}^L$ 到 $\varphi_{b\to a}^{L-1}$。

  end

end

$\varphi_{a\to b}=\varphi_{a\to b}^1$，$\varphi_{b\to a}=\varphi_{b\to a}^1$。

$A^*(p)=\frac{1}{n}\sum_{x\in N(p)}(B^*(\varphi_{a\to b}(x)))$，

$B(p)=\frac{1}{n}\sum_{x\in N(p)}(A(\varphi_{b\to a}(x)))$。

深度图像类比的方法在纹理迁移和颜色迁移有着出色的表现，但是计算时间较长。He 等人[错误!未找到引用源。]在 Liao 等人[错误!未找到引用源。]的工作基础上实现了一对一和一对多的图像颜色迁移，该方法主要针对图像的颜色进行处理，在满足局部约束和全局约束的情况下，通过类比的方法进行迭代优化，最终生成效果自然的新图像。

1.2 基于模型迭代的图像风格迁移方法

虽然基于图像迭代的方法可以产生效果出色的风格合成图像，但是存在计算效率低下的问题。而基于模型迭代的图像风格迁移方法使用了大量的图像来训练可生成风格化图像的生成模型，很大程度上解决了图像风格迁移的计算效率低下的问题，并且可以与基于图像迭代的方法进行组合。目前，应用市场上的应用程序主要使用基于模型迭代的方法。以下将对基于生成模型和基于图像重构解码器这两类具有代表性的方法展开详细探讨。

1.2.1 基于生成模型方法

Johnson 等人[错误!未找到引用源。]最早提出了迭代优化生成模型的图像风格迁移方法，亦称作快速风格迁移，如图 2 所示。这种方法建立在 Gatys 等人算法的基础上，使用感知损失函数为某种特定风格训练出一个生成模型。与之前训练生成模型时使用逐个像素比较的损失函数相比，感知损失函数对预训练 VGG 模型提取的高层抽象特征表示进行平方求差，这部分的问题求解与 Gatys 等人算法是一致的。具体而言，Johnson 等人的方法使用残差网络[错误!未找到引用源。]作为生成模型的基本组件，训练数据使用 COCO 数据集[错误!未找到引用源。]，感知损失函数可以表示如下：

$$\begin{aligned}y^*=\arg\min_{f(x)}&\lambda_c L_c(f(x),y_c)\\&+\lambda_s L_c(f(x),y_s)\\&+\lambda_{TV}(f(x))\end{aligned}\tag{12}$$

其中：λ_c、λ_s 和 λ_{TV} 分别表示内容损失函数的权重系数、风格损失函数的权重习俗和图像平滑函数的权重系数，x 表示网络输入，$f(x)$ 表示生成模型函数，y_c 表示预训练 VGG 模型从训练图像中提取的内容特征表示，y_s 表示预训练 VGG 模型从风格图像中提取的风格特征表示。

Johnson 等人的工作[错误!未找到引用源。]为如何提高图像风格迁移效

率提供了一个很好的启发。此外，Ulyanov 等人的工作错误!未找到引用源。也采用了类似的网络架构进行，并通过实验表明了在生成模型训练过程中，使用实例归一化错误!未找到引用源。替代批量归一化错误!未找到引用源。可以显著提高生成图像的质量。Wang 等人错误!未找到引用源。提出了一种多模式卷积神经网络，该方法考虑了颜色和亮度通道的特征表示，以多尺度分级地执行风格化，有效解决了纹理比例适配的问题，并在高分辨率图像上产生了可观的图像生成效果。Zhang 等人错误!未找到引用源。构建了一个可以训练多种风格的生成模型，实现了多风格的快速风格迁移。Huang 等人错误!未找到引用源。提出了一种自适应实例归一化的方法，解决了生成模型训练需要进行风格化预定义的问题。

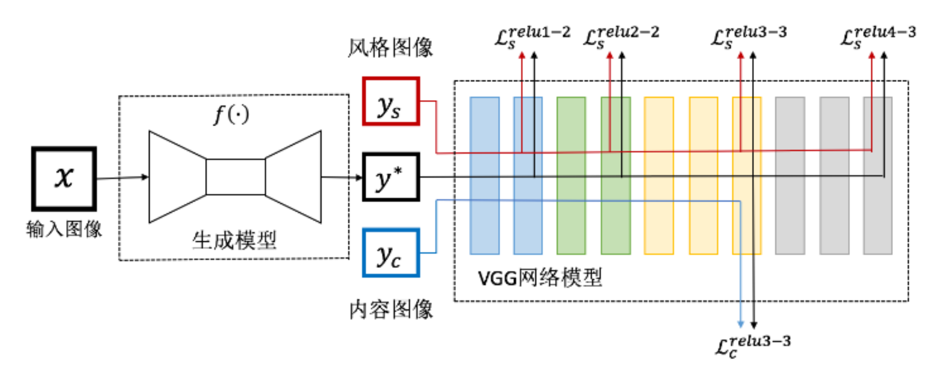


图 2　快速风格迁移架构

此外，生成对抗网络错误!未找到引用源。在图像风格迁移方面的应用同样有着出色的效果。Li 等人错误!未找到引用源。将马尔可夫随机场与生成对抗网络相结合，采用对抗训练的方式来训练生成模型，使得生成图像具有很好真实感。随后提出的无监督的生成对抗网络，如 CycleGAN 错误!未找到引用源。，DiscoGAN 错误!未找到引用源。和 DualGAN 错误!未找到引用源。，其中，CycleGAN 基于循环一致性，DiscoGAN 和 DualGAN 基于机器翻译的对偶学习思想错误!未找到引用源。。这些优秀的对抗训练模型突破了训练数据需要成对的限制，成功实现了无监督的迁移学习，并且在网络架构和具体实现上基本一致。本文以通用性较好的 CycleGAN 模型为例进行讲解。Zhu 等人错误!未找到引用源。提出的 CycleGAN 模型包含两个生成式模型 G 和 F，以及两个判别式模型 D_X 和 D_Y，并且使用了循环一致性作为总损失函数的约束条件。在 CycleGAN 模型中，总损失函数包含两个部分，即对抗损失函数和循环一致性损失函数，其中，对抗损失函数包含向前映射损失和向后映射损失两部分。具体而言，给定数据集 X 和 Y，其中 $x \in X$，$y \in Y$，向前映射损失函数表示如下：

$$L_{GAN}(G,D_Y,X,Y)=E_{y\sim p_{data}(y)}[\log D_Y(y)]+E_{x\sim p_{data}(x)}[\log(1-D_Y(G(x)))] \tag{13}$$

同理，向后映射损失函数表示如下：

$$L_{GAN}(F,D_X,Y,X)=E_{x\sim p_{data}(x)}[\log D_X(x)]+E_{y\sim p_{data}(y)}[\log(1-D_X(F(y)))] \tag{14}$$

而循环一致性损失函数是为了使生成式模型 G 和 F 之间保持一致性，可表示如下：

$$L_{cyc}(G,F)=E_{x\sim p_{data}(x)}[\|F(G(x))-x\|_1]+E_{y\sim p_{data}(y)}[\|G(F(y))-y\|_1] \tag{15}$$

最后，CycleGAN 的总损失函数表示如下：

$$L(G,F,D_X,D_Y)=L_{GAN}(G,D_Y,X,Y)+L_{GAN}(G,D_Y,X,Y)+\lambda L_{cyc}(G,F) \tag{16}$$

其中：λ 表示向前映射和向后映射这两个目标函数的相对重要性的平衡参数。最终，CycleGAN 模型的对抗训练的优化目标表示如下：

$$G^*,F^*=\arg\min_{F,G}\max_{D_X,D_Y} L(G,F,D_X,D_Y) \tag{17}$$

其中：G^* 和 F^* 分别表示最终所求得的近似最优的两个生成式模型。

但是，目前的生成对抗网络在模型训练方面相当不稳定，而判别模型的设定使得指向明确的图像风格迁移方法实现起来变得困难。此外，生成对抗网络是根据图像散度分布的迭代优化来进行对抗训练，而不是根据图像的内容、纹理和颜色，因此，使用生成对抗网络进行图像风格迁移的过程显得难以控制。

1.2.2 基于图像重构解码器方法

基于图像迭代存在着参数调整和效率低下的两个弊端，而快速风格迁移虽然缓解了效率低下的问题，但只能针对特定风格进行模型训练，并且仍然无法避免参数调整的问题。为了克服这些问题，Li 等人错误!未找到引用源。提出了一种基于图像重构解码器的图像风格迁移算法，该算法不再需要针对特定风格进行模型训练，避免了参数调整的问题。

该算法使用多层次的风格化策略，如图 3 所示，首先使用预训练 VGG 模型作为编码器，将其权重固定并用于训练解码器网络，以便将 VGG 特征反转为原始图像，其中解码器被设计成与编码器对称。解码器使用像素重建损失和特征损失作为重建图像的约束条件，其损失函数表示如下：

$$L=\|I_{output}-I_{input}\|_2^2+\lambda\|\Phi(I_{output})-\Phi(I_{input})\|_2^2 \tag{18}$$

其中：I_{output} 和 I_{input} 分别表示重构后的图像输出和输入图像，训练数据使用的是 COCO 数据集错误!未找到引用源。，Φ 表示使用预训练 VGG 模型编码器提取的图像特征表示，λ 表示像素重建损失和特征损失的平衡权重系数。

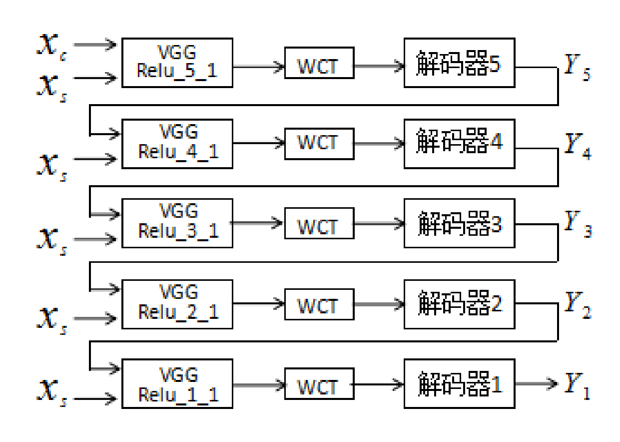


图 3　多层次风格化结构

当完成对应层的解码器训练后，在编码器 $En(\cdot)$ 和解码器 $De(\cdot)$ 之间设置对应的投影函数，通过白化和着色转换（whitening and coloring transform, WCT）进行程式化的图像重构。具体来说，给定内容图像 x_c 和风格图像 x_s，在预训练 VGG 模型中分别提取它们在特定层中的抽象特征表示并通过白化与着色转换分别求得内容图像和风格图像的矢量化特征 $H_c = En(x_c)$ 和 $H_s = De(x_s)$，然后求得对应层的风格化编码结果 H_{cs}，可以如下表示：

$$H_{cs} = P_s P_c H_c \tag{19}$$

其中：$P_c = E_c \frac{1}{\sqrt{D_c}} E_c^T$，$P_s = E_s \frac{1}{\sqrt{D_s}} E_s^T$，$D_c$ 和 D_s 分别是协方差矩阵 $H_c H_c^T$ 和 $H_s H_s^T$ 的对角矩阵，矩阵 E_c 和 E_s 分别是协方差矩阵 $H_c H_c^T$ 和 $H_s H_s^T$ 的正交矩阵。最后，使用训练好的解码器 $De(\cdot)$ 对风格化编码 H_{cs} 进行解码，可获得对应层的合成图像 $Y = De(H_{cs})$。为了获得更好的效果，Li 等人错误!未找到引用源。进一步改善了文献**错误!未找到引用源。**中的图像重构编码器的结构，并增加了图像局部平滑错误!未找到引用源。错误!未找到引用源。的后处理，实现了照片般逼真的快速风格迁移，与 Luan 等人错误!未找到引用源。提出的深度照片风格迁移方法的效果基本一致。

2　应用分析

随着基于深度学习的图像风格迁移在算法和理论方面的不断改进，图像风格迁移的效果得到很大的提升，具有广阔的商业应用前景。目前，基于深度学习的图像风格迁移的应用主要有以下三个方向。

a)图像处理。目前，在社交网络上流传的图像大多经过软件处理，其中图像美化是一个热门应用技术。传统的图像处理技术只能对图像进行模式较为固定的处理，而基于神经网络的图像风格迁移的出现为图像风格设计带来了更多的想象空间。Chen 等人错误!未找到引用源。提出了一种内容感知的风格迁移方法，同时该方法可有效地应用于图像修复。Zhang 等人错误!未找到引用源。提出了一种给漫画草图上色的方法。而 Prisma 是第一款免费提供基于深度学习的图像风格迁移服务的移动应用程序，能够在短短几秒内将用户的照片变成高质量的艺术画作。随后陆续出现了一些收费的图像风格迁移应用软件，产生了一定的商业价值。在这些应用程序的帮助下，人们可以轻松地创造出属于自己风格的艺术作品，而不需要具备特殊的专业技能。

b)视频处理。在影视娱乐产业中，如电影、电视、动漫等等，影视特效技术随处可见。但是，影视特效技术的创作不仅需要特殊的专业技能，还需要大量的手工劳动，如果可以使用更多的人工智能技术，将可以大幅度地降低制作成本，而图像风格迁移是一个可考虑的解决方案。例如，Anderson 等人错误!未找到引用源。使用光流和深度神经网络来进行电影风格化。Ruder 等人错误!未找到引用源。引入了时间一致性损失函数来提高视频风格化后的帧之间的连贯性。Chen 等人错误!未找到引用源。构建了一个具有时间相关性的网络模型，它可以包含多种风格，并且能够实时风格化在线视频。Joshi 等人错误!未找到引用源。深入研究图像风格迁移中更高级的参数空间，并找到一组有效的组件对电影场景进行印象派风格化。

c)风格设计的辅助工具。图像风格迁移可以充当有用的辅助工具，例如艺术绘画创作、建筑艺术设计、服装艺术设计、游戏场景设计等等。虽然目前还没有相关的参考文献或成功的应用案例，但这很有可能会成为未来的研究热点。

从目前的研究进展来看，基于深度学习的图像风格迁移正快速发展，如何提高算法效率和图像质量仍然有很大的研究空间，其潜在的商业价值有待进一步挖掘。

3　存在问题与研究方向

基于深度学习的图像风格迁移的算法已经取得了显著的效果，但仍然存在一些尚待解决的问题。本章总结了目前主要存在的问题，并提出了一些建议。

a)参数调整。为了获得令人满意的结果，基于图像迭代方法和基于模型迭代方法都需要手动调参，特别是基于模型迭代方法，每次调整参数后都要重新对模型进行训练。虽然基于图像重构编码器方法能缓解了参数调整的问题，并且不需要为不同样式单独训练模型，但是图像重构解码器的训练过程较为繁琐，图像生成效果也并不太理想。而局部平滑处理能改善基于图像重构解码器的方法，但会使得风格化图像的纹理消失，最终的效果与图像颜色迁移错误!未找到引用源。几乎相似。因此，找到一种既简单可控，又能保证图像质量的方法是下一步的重要研究方向。如果在不考虑模型存储容量大小的情况下，进一步改善基于图像重构编码器方法的图像生成质量是一个非常值得考虑的研究方向，因为该方法可以有效地避免参数调整的问题。

b)预训练模型的限制。Gatys 等人错误!未找到引用源。发现了使用预训练 VGG 模型可以提取图像的高层抽象特征，通过迭代优化进而可以实现图像风格迁移。到目前为止，大多数基于深度学习的图像风格迁移方法都使用预训练 VGG 模型进行图像特征提取。虽然预训练 VGG 是一个优秀的卷积神经网络模型，在特征提取方面表现出色，但它是一个重量级的模型，存在体积庞大和计算量巨大的问题，并且最初并不是专门为图像风格迁移而设计的。因此，摆脱对预先练 VGG 模型的依赖或设计出

更精小、更有效的特征提取器是推动基于深度学习的图像风格迁移进一步发展的重要途径。生成对抗网络或许可以解决预训练模型的限制问题，其逼真的图像生成效果有利于提高生成图像的质量，而且基于散度分布的优化方式与基于图像迭代的相关方法具有一定的相似度，并且对抗训练的方式在获取新特征的应用场景中具有很好的效果。

c)迁移学习理论的完善。图像风格迁移是迁移学习中的一个典型用例。目前，基于深度学习的迁移学习方法仍然处于起步阶段，需要有更完备的数学方法和理论指导。迁移学习理论的完善对基于深度学习的图像风格迁移的进一步发展具有非常重要的意义。而通用模型的相关研究工作错误!未找到引用源。错误!未找到引用源。提出了设计通用性强的神经网络模型，以提高模型的迁移学习能力，对图像风格迁移进一步的发展有着重要的指导作用。

d)预处理和后处理的方法。为了使得最终结果更符合实际应用，可以使用一些预处理和后处理的方法，如图像语义分割错误!未找到引用源。、图像融合错误!未找到引用源。、图像颜色迁移错误!未找到引用源。错误!未找到引用源。、图像平滑处理错误!未找到引用源。错误!未找到引用源。等。这些预处理和后处理的方法对提升图像风格迁移的效果有重要作用，例如，Castillo 等人错误!未找到引用源。结合图像语义分割的方法对图像中特定的对象进行风格迁移。Li 等人错误!未找到引用源。的工作结合图像融合技术为用户提供了友好的交互。Gatys 等人错误!未找到引用源。使用图像颜色迁移方法来实现风格化图像的颜色控制。Li 等人错误!未找到引用源。对风格化图像进行后期的图像局部平滑处理以获得照片般的效果，如图 4 所示。因此，结合有效的预处理和后处理方法是提高风格迁移结果的重要手段。

内容图　　风格图　　Gatys 等人方法　　局部平滑处理



图 4　图像后处理效果对比

4　结束语

本文对基于深度学习的图像风格迁移进行了详细的介绍，对其应用前景、存在问题和发展方向展开了深入的探讨和分析。虽然目前已经有了成功的应用案例，但是距离广泛的商业应用还有较大的距离，需要更进一步的研究和完善。总体而言，基于深度学习的图像风格迁移是一个具有挑战性的新兴课题，它不仅受到了学术界的广泛关注，而且在工业界中也有很大的需求，具有重要的研究意义和广阔的应用前景。

参考文献：

[1] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer [C]// Proc of the 28th annual conference on Computer graphics and interactive techniques. New York: ACM Press, 2001: 341-346.

[2] Hertzmann A, Jacobs C E, Oliver N, *et al.* Image analogies [C]// Proc of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 327-340.

[3] 张海嵩，尹小勤，于金辉. 实时绘制 3D 中国画效果 [J]. 计算机辅助设计与图形学学报, 2004, 16 (11): 1485-1489. (Zhang Haisong, Yin Xiaoqin, Yu Jinhui. Real-time Rendering of 3D Chinese Painting Effects [J]. Journal of Computer-Aided Design & Computer Graphics, 2004, 16 (11): 1485-1489.)

[4] 钱小燕，肖亮，吴慧中. 快速风格迁移 [J]. 计算机工程, 2006, 32 (21): 15-17. (Qian Xiaoyan, Xiao Liang, Wu Huizhong. Fast Sytle Transfer [J]. Computer Engineering, 2006, 32 (21): 15-17.)

[5] 毛勇华，桂小林，李前，等. 深度学习应用技术研究 [J]. 计算机应用研究, 2016, 33 (11): 3201-3205. (Mao Yonghua, Gui Xiaolin, Li Qian, *et al.* Study on application technology of deep learning [J]. Application Research of Computers, 2016, 33 (11): 3201-3205.)

[6] 刘建伟，刘媛，罗雄麟. 深度学习研究进展 [J]. 计算机应用研究, 2014, 31 (7): 1921-1930. (Liu Jianwei, Liu Yuan, Luo Xionglin. Research and development on deep learning [J]. Application Research of Computers, 2014, 31 (7): 1921-1930.)

[7] 孙志军，薛磊，许阳明，等. 深度学习研究综述 [J]. 计算机应用研究, 2012, 29 (8): 2806-2810. (Sun Zhijun, Xue Lei, Xu Yangming, *et al.* Overview of deep learning [J]. Application Research of Computers, 2012, 29 (8): 2806-2810.)

[8] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style [J]. arXiv preprint arXiv: 1508. 06576, 2015.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.

[10] Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks [J]. arXiv preprint arXiv: 1505. 07376, 2015.

[11] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional

neural networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society Press, 2016: 2414-2423.

[12] Gatys L A, Bethge M, Hertzmann A, *et al.* Preserving color in neural artistic style transfer [J]. arXiv preprint arXiv: 1606. 05897, 2016.

[13] Gatys L A, Ecker A S, Bethge M, *et al.* Controlling perceptual factors in neural style transfer [J]. arXiv preprint arXiv: 1611. 07865, 2016.

[14] Li Yanghao, Wang Naiyan, Liu Jiaying, *et al.* Demystifying neural style transfer [J]. arXiv preprint arXiv: 1701. 01036, 2017.

[15] Li Chuan, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2016: 2479-2486.

[16] Liao Jing, Yao Yuan, Yuan Lu, *et al.* Visual attribute transfer through deep image analogy [J]. arXiv preprint arXiv: 1705. 01088, 2017.

[17] Johnson J, Alahi A, Li Feifei. Perceptual losses for real-time style transfer and super-resolution [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2016: 694-711.

[18] Huang Haozhi, Wang Hao, Luo Wenhan, *et al.* Real-time neural style transfer for videos [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 7044-7052.

[19] Wang Xin, Oxholm G, Zhang Da, *et al.* Multimodal transfer: a hierarchical deep convolutional neural network for fast artistic style transfer [J]. arXiv preprint arXiv: 1612. 01895, 2016.

[20] Li Yijun, Fang Chen, Yang Jimei, *et al.* Universal style transfer via feature transforms [J]. arXiv preprint arXiv: 1705. 08086, 2017.

[21] Li Yijun, Liu Mingyu, Li Xueting, *et al.* A closed-form solution to photorealistic image stylization [J]. arXiv preprint arXiv: 1802. 06474, 2018.

[22] Nikulin Y, Novak R. Exploring the neural algorithm of artistic style [J]. arXiv preprint arXiv: 1602. 07188, 2016.

[23] Novak R, Nikulin Y. Improving the neural algorithm of artistic style [J]. arXiv preprint arXiv: 1605. 04603, 2016.

[24] Risser E, Wilmot P, Barnes C. Stable and controllable neural texture synthesis and style transfer using histogram losses [J]. arXiv preprint arXiv: 1701. 08893, 2017.

[25] Yin Rujie. Content aware neural style transfer [J]. arXiv preprint arXiv: 1601. 04568, 2016.

[26] Efros A A, Leung T K. Texture synthesis by non-parametric sampling [C]// Proc of the 7th IEEE International Conference. Piscataway, NJ: IEEE Press, 1999, 2: 1033-1038.

[27] Champandard A J. Semantic style transfer and turning two-bit doodles into fine artworks [J]. arXiv preprint arXiv: 1603. 01768, 2016.

[28] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2016: 770-778.

[29] Lin T Y, Maire M, Belongie S, *et al.* Microsoft coco: Common objects in context [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2014: 740-755.

[30] Ulyanov D, Lebedev V, Vedaldi A, *et al.* Texture networks: feed-forward synthesis of textures and stylized images [J]. arXiv preprint arXiv: 1603. 03417, 2016.

[31] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization [J]. arXiv preprint arXiv: 1607. 08022, 2016.

[32] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. arXiv preprint arXiv: 1502. 03167, 2015.

[33] Zhang Hang, Dana K. Multi-style generative network for real-time transfer [J]. arXiv preprint arXiv: 1703. 06953, 2017.

[34] Huang Xun, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [J]. arXiv preprint arXiv: 1703. 06868, 2017.

[35] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets [C]// Proc of International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.

[36] Li Chuan, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2016: 702-716.

[37] Zhu Junyan, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks [J]. arXiv preprint arXiv: 1703. 10593, 2017.

[38] Kim T, Cha M, Kim H, *et al.* Learning to discover cross-domain relations with generative adversarial networks [J]. arXiv preprint arXiv: 1703. 05192, 2017.

[39] Yi Zili, Zhang Hao, Tan Ping, *et al.* DualGAN: unsupervised dual learning for image-to-image translation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 2849-2857.

[40] Xia Yingce, He Di, Qin Tao, *et al.* Dual learning for machine translation [J]. arXiv preprint arXiv: 1611. 00179, 2016.

[41] Zhou Dengyong, Weston J, Gretton A, *et al.* Ranking on data manifolds [C]// Advances in Neural Information Processing Systems. 2003: 169-176.

[42] Yang Chuan, Zhang Lihe, Lu Huchuan, *et al.* Saliency Detection via Graph-Based Manifold Ranking [C]// Proc of Computer Vision and Pattern Recognition Conference. Piscataway, NJ: IEEE Press, 2013: 3166-3173.

[43] Luan F, Paris S, Shechtman E, *et al.* Deep photo style transfer [J]. arXiv preprint arXiv: 1703. 07511, 2017.

[44] Chen Yilei, Hsu C T. Towards deep style transfer: a content-aware perspective [C]// Proc of British Machine Vision Conference. 2016: 8. 1-8. 11.

[45] Anderson A G, Berg C P, Mossing D P, *et al.* DeepMovie: using optical flow and deep neural networks to stylize movies [J]. arXiv preprint arXiv: 1605. 08153, 2016.

[46] Zhang Lvmin, Ji Yi, Lin Xin. Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier GAN [J]. arXiv preprint arXiv: 1706. 03319, 2017.

[47] Ruder M, Dosovitskiy A, Brox T. Artistic style transfer for videos [J]. arXiv preprint arXiv: 1604. 08610, 2016.

[48] Chen Dongdong, Liao Jing, Yuan Lu, *et al.* Coherent online video style transfer [J]. arXiv preprint arXiv: 1703. 09211, 2017.

[49] Joshi B, Stewart K, Shapiro D. Bringing impressionism to life with neural style transfer in come swim [J]. arXiv preprint arXiv: 1701. 04928, 2017.

[50] He Mingming, Liao Jing, Yuan Lu, *et al.* Neural color transfer between images [J]. arXiv preprint arXiv: 1710. 0756, 2017.

[51] Castillo C, De S, Han X, *et al.* Son of Zorn's Lemma: targeted style transfer using instance-aware semantic segmentation [J]. arXiv preprint arXiv: 1701. 02357, 2017.

[52] Garcia-Garcia A, Orts-Escolano S, Oprea S, *et al.* A review on deep learning techniques applied to semantic segmentation [J]. arXiv preprint arXiv: 1704. 06857, 2017.

[53] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2008, 30 (2): 228-242.

[54] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need [J]. arXiv preprint arXiv: 1706. 03762, 2017.

[55] Kaiser L, Gomez A N, Shazeer N, *et al.* One model to learn them all [J]. arXiv preprint arXiv: 1706. 05137v1, 2017.